

Rappresentazione delle informazioni come configurazione di bit

Informazione

Una definizione di Informazione potrebbe essere: tutto ciò che riduce incertezza. (**Es:** Una foto riduce l'incertezza di chi la osserva riguardo alla scena raffigurata. L'espressione del volto di un interlocutore riduce l'incertezza sul suo stato d'animo. Un fruscio riduce l'incertezza sul luogo in cui cercare un fagiano. Ogni parola riduce l'incertezza sul significato di una frase.)

D'altra parte non bisogna confondere l'informazione con un dato. Infatti quando parliamo di dato intendiamo informazione codificata. L'informazione in un dato viene quindi codificata in un linguaggio convenzionale per poter essere rappresentata, scambiata, memorizzata ed elaborata. La codifica delle informazioni può rendere omogenea la rappresentazione e l'elaborazione di informazioni di natura diversa (suoni, immagini, testi, numeri).

Informazione oggi

Oggi quando si parla di informazione non si può fare a meno di parlare di informatica. Sappiamo che *informatica* significa "informazione automatica". Più in particolare diciamo che l'informatica è quella disciplina che studia l'elaborazione automatica delle informazioni. Per automatico si intende tutto ciò che è autonomo, indipendente e che compie un compito prestabilito senza l'intervento umano. Quindi quando si parla di **automazione** si fa riferimento alla realizzazione di strumenti per la soluzione di problemi o l'esecuzione di processi in maniera indipendente dall'attività umana.

Normalmente un qualche processo viene automatizzato quando il numero di volte che esso deve essere eseguito è talmente grande da rendere conveniente la progettazione e la costruzione di un sistema automatico che lo risolva. Gli stessi problemi che vengono trattati in maniera automatica possono essere trattati dall'uomo, anzi, più precisamente sono stati originariamente trattati dall'uomo.

Es:

Problema originale (non ripetitivo)
Soluzione manuale

Problema sistematico e ripetitivo
Soluzione automatica

Determinare il diametro di una pallina
diametro

Calibro

Classificare un insieme di palline in base al

Griglie forate

Scrivere una lettera
Carta e penna

Riprodurre un testo in migliaia di copie
Stampa tipografica

Regolare il traffico in caso di emergenza
Vigile urbano

Regolare il traffico ad ogni incrocio
Semaforo

Riempire un contenitore d'acqua
d'acqua

Rubinetto manuale

Riempire ripetutamente un contenitore con la stessa quantità

Rubinetto a galleggiante

Computer (elaboratore elettronico)

Un elaboratore elettronico è una macchina (elettronica) programmabile per l'elaborazione automatica di dati. Un elaboratore è costituito da una parte così detta *Hardware* (componenti tangibili e permanenti di un computer). E da una parte *software* (insieme dei programmi, componenti intangibili e temporanei, che controllano il funzionamento dell'hardware).

L'hardware è in grado di eseguire un insieme finito di *operazioni* elementari e di interpretare *istruzioni* in un linguaggio predefinito. Il software specifica sequenze di istruzioni da eseguire per arrivare alla soluzione di un problema.

Programmazione

Un elaboratore è *programmabile* se è in grado di svolgere compiti diversi in base ad un *programma*. Un elaboratore programmabile può essere utilizzato ripetutamente per risolvere problemi diversi o *istanze* diverse dello stesso problema. In molti casi, l'automazione di un processo è conveniente solo se si può realizzare con elaboratori programmabili già progettati e sviluppati per altre applicazioni. In tal caso solo il programma deve essere riprogettato. Il processo denominato *codifica* consente di utilizzare lo stesso elaboratore per manipolare informazioni eterogenee. Ogni informazione può (a meno di un'approssimazione) essere rappresentata come una sequenza di due soli simboli (**Es:** 0 e 1) che nel campo informatico vengono chiamati *BIT*.

BIT

La parola BIT nasce dalla contrazione delle parole inglesi *Binary Digit* e letteralmente BIT significa "cifra binaria". Cioè una cifra che può assumere solo due valori (di solito si parla di 0 e 1). La rappresentazione fisica di un bit si può ottenere da un qualunque sistema in grado di trovarsi in due stati diversi mutuamente esclusivi. Ad esempio un bit può essere rappresentato da un interruttore che può trovarsi nei due stati acceso/spento. L'interruttore infatti si trova alla base del funzionamento dei circuiti logici che costituiscono le unità di elaborazione dei microprocessori. Allo stesso modo un bit può essere rappresentato da un condensatore che si trova nei due stati carico/scarico. Tutte le memorie principali così dette *dinamiche* si basano su questo principio, memorizzando i singoli bit in piccolissimi condensatori in modo tale da rappresentare un 1 se il condensatore è carico e uno 0 se questo è scarico. D'altra parte, nelle memorie di massa non volatili come i *dischi fissi*, i bit vengono memorizzati attraverso particelle magnetiche orientabili. In questo modo per rappresentare un bit di valore 1 si orienta una particella in una direzione e per rappresentare uno 0 si orienta nel verso opposto.

In fine, anche un sistema che non abbia nulla a che fare con un computer può rappresentare un bit. Ad esempio, una bandiera che si trovi nei due stati alzata/abbassata.

Codifica delle informazioni

La codifica delle informazioni in termini di bit è l'operazione che consente di utilizzare i calcolatori elettronici per elaborare dati di qualsiasi natura. Abbiamo già detto che un bit può rappresentare solamente due diverse informazioni (1 o 0, acceso o spento, ecc.). Per codificare insieme con un numero maggiore di informazioni/stati usiamo *stringhe* di bit di lunghezza variabile. Una stringa è una sequenza di n bit e viene anche detta parola dall'inglese *word*. In generale vale la regola che con n bit si possono formare 2^n configurazioni diverse e quindi si possono rappresentare 2^n informazioni diverse. La codifica è una convenzione che associa un significato ad ogni configurazione di bit.

I calcolatori sono oggetti finiti che elaborano e memorizzano un numero finito di bit. Un numero finito di bit permette di codificare un numero finito di informazioni diverse. Una codifica esatta a n bit è possibile solo quando l'insieme delle informazioni da codificare è *finito* e di dimensione inferiore a 2^n . Se l'insieme ha cardinalità maggiore di 2^n o è infinito, se ne può dare solo una rappresentazione *approssimata* o *parziale*. Le operazioni da compiere per ridurre la quantità di informazione da codificare sono le seguenti:

- *limitazione*, rappresentazione di un sottoinsieme di informazioni di dimensione 2^n (gli elementi del sottoinsieme sono rappresentati esattamente, gli altri non sono rappresentati);

- *partizionamento*, suddivisione dell'insieme in 2^n sottoinsiemi ad ognuno dei quali si associa una diversa configurazione di bit (la codifica permette di discriminare sottoinsiemi diversi, ma non elementi dello stesso sottoinsieme).

Un metodo di partizionamento comunemente utilizzato per insiemi numerici è detto *discretizzazione* o *quantizzazione* e consiste nella suddivisione regolare in intervalli di uguale ampiezza. Insiemi *infiniti* e *discreti* devono essere limitati per poter essere codificati. Insiemi *limitati* e *continui* devono essere partizionati. Insiemi *illimitati* e *continui* devono essere sia limitati che partizionati. La perdita di informazione associata alle operazioni di limitazione e discretizzazione rende la codifica approssimata.

Codifica del Testo

Un testo è una sequenza di caratteri alfabetici, separatori e caratteri speciali come punteggiatura, accentuazioni, ecc. La codifica *ASCII* (*American Standard Code for Information Interchange*) prevede l'uso di 128 caratteri diversi. Ogni carattere è associato ad una diversa configurazione di 7 bit. Estensioni non standard del codice ASCII utilizzano 8 bit (1 byte) per carattere, portando a 256 il numero di caratteri disponibili. Un testo è rappresentato dalla sequenza di byte associati ai caratteri che lo compongono, nell'ordine in cui essi compaiono. Un testo di 1000 caratteri richiede 1000 byte (1 Kb) per essere rappresentato.

Es: Il testo "ciao" richiede 4 byte per essere codificato secondo il codice ASCII a 8 bit. Il testo "ciao a tutti" ne richiede 12; il testo "perche' no" ne richiede 10, mentre il testo "perché no" ne richiede 9. Il testo "uno" richiede 3 byte.

Codifica dei numeri

Per la rappresentazione dei numeri viene comunemente usata la *notazione posizionale*. Nella notazione posizionale ogni simbolo assume un diverso valore a seconda della posizione occupata nel *grafema*. Per intenderci nel numero "12" il valore nel simbolo 1 è "dieci" in quanto occupa la posizione delle decine. Per la rappresentazione dei numeri è particolarmente diffusa la *codifica binaria*, che fornisce una diretta trasposizione in un alfabeto a due valori della notazione *decimale* da noi comunemente utilizzata. Il numero di simboli dell'alfabeto è detto *base* di numerazione. La codifica binaria e quella decimale utilizzano basi diverse (2 e 10, rispettivamente) ma sono entrambe *notazioni posizionali*, in quanto permettono di rappresentare ogni numero come sequenza di cifre del tipo:

$$C_n C_{n-1} \dots C_1 C_0 \cdot C_{-1} C_{-2} \dots C_{-m}$$

Il valore (v) del numero rappresentato è la somma pesata delle cifre, dove il peso di ogni cifra dipende dalla base di numerazione (B) e dalla posizione della cifra:

$$v = C_n B^n + C_{n-1} B^{n-1} + \dots + C_1 B + C_0 + C_{-1} B^{-1} + C_{-2} B^{-2} + \dots + C_{-m} B^{-m}$$

Es: $(101.1)_{(2)} = 1 * 2^2 + 0 * 2^1 + 1 * 2^0 + 1 * 2^{-1} = (5.5)_{(10)}$.

Numeri interi. I numeri interi sono un insieme discreto illimitato. Per poter essere codificati devono essere limitati. In genere ci si restringe ad un sottoinsieme simmetrico rispetto allo 0. Avendo a disposizione n bit, se ne usano 1 per rappresentare il *segno* e i restanti $a = n-1$ per rappresentare il modulo. Il massimo numero rappresentabile è (in modulo) $2^a - 1$. Se il risultato di un'operazione eccede il modulo $2^a - 1$ non può essere codificato e il calcolatore restituisce un messaggio di *overflow*.

Numeri reali. I numeri reali sono un insieme continuo e illimitato. Per poterli rappresentare occorre limitarli (in modo simmetrico rispetto allo 0) e partizionarli.

Rappresentazione in virgola fissa. Degli n bit della parola, 1 rappresenta il segno, a rappresentano le cifre prima della virgola e b le cifre dopo la virgola

Rappresentazione in virgola mobile (floating-point). Il numero da rappresentare viene espresso nella forma

$$s \ O.M \ B^{se \ E}$$

dove s rappresenta il segno, M la *mantissa*, E l'esponente, se il segno dell'esponente e B la base di numerazione. Non tutti i numeri hanno rappresentazione univoca. Questo comporta perdita di efficienza e difficoltà di confronto. Quando esistono rappresentazioni equivalenti dello stesso numero, tra queste si definisce *normale* quella con l'esponente più piccolo. Da una rappresentazione non normale a quella normale equivalente si passa traslando verso sinistra le cifre della mantissa e sottraendo 1 all'esponente.

Codifica delle immagini

Le immagini sono informazioni continue in tre dimensioni: due spaziali ed una colorimetrica. Per codificarle occorre operare tre discretizzazioni. Le due discretizzazioni spaziali riducono l'immagine ad una matrice di punti colorati, detti *pixel*. La terza discretizzazione limita l'insieme di colori che ogni pixel può assumere. Il modo più elementare per rappresentare un'immagine consiste nel rappresentare il colore di ogni pixel. Il numero di bit necessari a rappresentare ogni pixel dipende dal numero di diversi colori disponibili. Un byte consente di rappresentare 256 colori. Un'immagine a 256 colori è una sequenza di byte.

Un'immagine di 100X100 pixel a 256 colori richiede 10000 byte (10 Kb) per essere rappresentata. Un'immagine a due soli colori (ad esempio bianco e nero) richiede un solo bit per pixel. Le immagini codificate secondo questa tecnica vengono dette anche immagini *raster* o immagini *bitmap*. Questo tipo di immagini vengono solitamente usate per fotografie, scansioni e in genere per tutte le immagini che prevedono sottili gradazioni nei colori e nelle forme.

Immagini vettoriali: la grafica vettoriale scompone l'immagine in gruppi logici di componenti (linee, cerchi, rettangoli, ecc.) e memorizza le forme in termini di coordinate e colori dei vari elementi geometrici che le compongono. In definitiva un file che memorizza un'immagine di tipo vettoriale è costituito da un'insieme di coordinate e dalle informazioni sul colore di ciascuna forma. Nel momento della visualizzazione i punti descritti dalle coordinate vengono disegnati e colorati a dovere. La grafica vettoriale viene comunemente usata nel disegno animato, nella grafica lineare e negli strumenti CAD. Un grande vantaggio delle immagini vettoriali rispetto a quelle raster/bitmap sta nella minor dimensione del file che le memorizza e nella capacità di essere variate di dimensioni senza subire alcuna distorsione. Le immagini bitmap, invece, se ridimensionate rispetto alle dimensioni originali di acquisizione, hanno la tendenza a perdere di risoluzione risultando distorte o sfocate.

Filmati video (immagini in movimento). I filmati video possono essere considerati sequenze di fotogrammi. Abbiamo già visto come codificare immagini, e quindi sappiamo come codificare ogni fotogramma. Il numero di fotogrammi nell'unità di tempo determina la qualità del movimento. E' stato sperimentato che l'occhio umano non percepisce discontinuità nel movimento al di sotto del venticinquesimo di secondo. Pertanto, una frequenza di riproduzione dei fotogrammi superiore a 25Hz sarebbe superflua se lo scopo fosse la riproduzione di un video. D'altro canto potrebbe non esserlo se servisse a documentare un esperimento di brevissima durata o se il video dovesse essere rivisto al rallentatore.

Es: Filmato di 10 minuti, con risoluzione di 100x100 pixel a 256 colori: dimensione complessiva di $600 \times 25 \times 100 \times 100 \times 8 = 1.2 \text{Gbit}$.

Codifica di segnali analogici

Un segnale analogico è una grandezza fisica che varia nel tempo e che può assumere un insieme continuo (e quindi infinito) di valori. Un segnale digitale invece può assumere solo un insieme discreto (e generalmente finito) di valori. Un segnale analogico ha anche la proprietà di essere *tempo-continuo*: cioè il suo valore è significativo (e può variare) in qualsiasi istante di tempo. Diversamente, un segnale digitale viene anche detto segnale *tempo-discreto*: poiché il suo valore ha interesse solo in istanti di tempo prestabiliti, generalmente equidistanziati, mentre non ha alcun valore tra un istante ed un altro.

La rappresentazione più naturale di un segnale s è la sequenza dei suoi valori istantanei $s(t)$. Ci poniamo prima il problema di rappresentare un *valore istantaneo*, e poi quello di rappresentarne la *sequenza*. Il valore istantaneo di un segnale può essere rappresentato da un numero, e come tale può essere codificato utilizzando notazione in virgola fissa o in virgola mobile. Tali codifiche inducono limitazioni e partizionamenti sullo spazio dei possibili valori che il segnale può assumere (quantizzazione del segnale). Una volta codificato il valore istantaneo del segnale si ottiene la sequenza temporale ripetendo la codifica ad intervalli prestabiliti. In generale, se il segnale da codificare era tempo-discreto, la scelta naturale è quella di rappresentare il valore del segnale in tutti e soli gli istanti in cui esso è significativo. Per la codifica dei segnali analogici, che come abbiamo detto sono tempo-continui occorre operare una *discretizzazione* dell'intervallo temporale per ottenere una rappresentazione finita. L'operazione di discretizzazione più comune è il *campionamento*, che consiste nell'osservare il segnale tempo-continuo ad istanti di tempo prefissati (generalmente equidistanziati) in modo da ottenere un segnale tempodiscreto che assume lo stesso valore di quello originale nei punti di campionamento.

In definitiva, le operazioni di *campionamento* e *quantizzazione* trasformano un segnale analogico tempo-continuo in un segnale digitale tempo-discreto.

Codifica del suono

Il suono è un segnale analogico tempo-continuo. Ci sono diversi modi per descrivere la natura di tale segnale. Per fissare le idee pensiamo che il suono prodotto da un altoparlante è prodotto dalla vibrazione di una membrana. Descrivendo la posizione della membrana nel tempo (e quindi il suo spostamento) a tutti gli effetti descriviamo il suono. Assumiamo allora che il segnale analogico $s(t)$ esprima la posizione della membrana (rispetto ad un riferimento fissato) all'istante di tempo t . Tale segnale deve essere campionato e discretizzato per poter essere codificato in termini di bit. Per valutare la qualità della codifica possiamo pensare di usare il segnale quantizzato e tempo-discreto (quello cioè effettivamente rappresentato dalla nostra codifica) per muovere la membrana di un altoparlante. Se il campionamento è troppo rado e vengono usati pochi bit per codificare ogni valore istantaneo, la perdita di informazione degrada la qualità del suono (il suono riprodotto è sensibilmente diverso da quello originale).

Per quanto riguarda il campionamento, è dimostrato che il suono percepibile dall'orecchio umano viene riprodotto fedelmente se la frequenza di campionamento (il numero di campioni in un secondo) è non inferiore a 30KHz. Per rendere intelligibile il parlato è sufficiente una frequenza di campionamento di 8KHz. La quantizzazione introduce comunque una distorsione. Questa tecnica di codifica del suono è quella comunemente usata nella codifica di tipo wave che troviamo nei file .wav. Questa codifica tenta di ricostruire fedelmente la forma d'onda del suono campionandola adeguatamente. Il file che si ottiene è normalmente di grandi dimensioni.

Es: Telefono. La comunicazione telefonica deve riprodurre la voce umana ai due estremi del collegamento garantendo la comprensione del parlato. A tal fine lo standard prevede un campionamento a 8KHz ed una quantizzazione a 256 livelli (codificati con 8 bit). Il segnale digitale che ne risulta usa 64Kbit per codificare ogni secondo di conversazione. 10 minuti di conversazione richiedono 38.4Mbit.

La codifica di tipo MIDI, invece ha tra i formati audio lo stesso ruolo che il formato vettoriale ha tra i formati grafici. Infatti, il formato MIDI acronimo di *Musical Instrument Digital Interface*, non è una registrazione sonora ma contiene dei comandi da impartire ad un sintetizzatore sonoro (software o hardware) che riproduce il suono di uno strumento musicale. I file MIDI per questo motivo sono molto piccoli ed hanno il formato del tipo:

suona la nota x per un tempo y con lo strumento z.

La codifica MIDI, quindi riproduce un suono in maniera sintetica. Le nuove suonerie polifoniche dei telefoni cellulari di ultima generazione si basano su una codifica di tipo MIDI.